FEMS Microbiology Letters 237 (2004) 333–340

FEMS MICROBIOLOGY Letters

www.fems-microbiology.org

Aspergillus flavus expressed sequence tags for identification of genes with putative roles in aflatoxin contamination of crops

Jiujiang Yu [a,*], Catherine A. Whitelaw [b], William C. Nierman [b,c], Deepak Bhatnagar [a], Thomas E. Cleveland [a]

[a] *Food and Feed Safety Research, USDA/ARS, Southern Regional Research Center, 1100 Robert E. Lee Blvd., New Orleans, LA 70124, USA*
[b] *The Institute for Genomic Research, 9712 Medical Center Drive, Rockville, MD 20850, USA*
[c] *Department of Biochemistry and Molecular Biology, The George Washington University School of Medicine, 2300 Eye Street NW, Washington, DC 20037, USA*

Received 30 April 2004; received in revised form 16 June 2004; accepted 29 June 2004

First published online 14 July 2004

Abstract

Aflatoxins, produced primarily by *Aspergillus flavus* and *A. parasiticus*, are among the most toxic and carcinogenic naturally occurring compounds. In an attempt to identify genes potentially involved in aflatoxin contamination of crops, and to better understand the biology of *A. flavus*, a large scale sequencing of *A. flavus* expressed sequence tags (EST) was conducted. The 5′ ends of 26,110 cDNA clones from a normalized cDNA expression library were sequenced. After annotation, a total of 7218 unique ESTs in *A. flavus* were assembled into 3749 tentative concensus sequences and 3469 singleton sequences. The functional classifications of the genes or Gene Ontology (GO) terms were assigned to these ESTs. Genes potentially involved in the aflatoxin contamination process were identified in the ESTs sequenced. These include the aflatoxin biosynthetic pathway, signal transduction, global regulation, pathogenicity of the fungus, and stress response.

Keywords: *Aspergillus flavus*; Expressed sequence tags; Aflatoxin biosynthesis; Mycotoxins; Genetic regulation; Pathogenesis

1. Introduction

Aspergillus flavus produces the secondary metabolites aflatoxins B_1 and B_2 and other mycotoxins such as cyclopiazonic acid. *A. flavus* is the predominant species [1,2] responsible for aflatoxin contamination of crops prior to harvest or during storage. The acute toxicity of aflatoxins and the carcinogenic property of aflatoxins were established and recognized for over 40 years [3,4]. Due to the significant health and economic impacts of aflatoxin contamination, the chemistry, enzymology, and genetics of the aflatoxin biosynthetic pathway in *A. flavus* and *A. parasiticus* have been actively studied [5–7]. Genetic studies on aflatoxin biosynthesis in *A. flavus* and *A. parasiticus* led to the cloning of 25 clustered genes within a 70 kb DNA region responsible for the enzymatic conversions in the aflatoxin biosynthetic pathway [8,9]. Regulatory elements such as *aflR* [10,11] and *aflS* (*aflJ*) [12,13], nutritional and environmental factors [14,15], fungal developmental and sporulation [16–19] were also found to affect aflatoxin formation. In *A. flavus* there are eight chromosomes with an estimated genome size of about 33–36 Mbp that harbor an estimated 12,000 functional genes (reviewed in [7,20], Dr. Machida, personal communication). However, the global regulatory control of aflatoxin formation, the

* Corresponding author Tel.: +1-504-286-4405; fax: +1-504-286-4419.
E-mail address: jiuyu@srrc.ars.usda.gov (J. Yu).

0378-1097/$22.00
doi:10.1016/j.femsle.2004.06.054

genes controlling *aflR* and *aflS* (*aflJ*) expression, the process of signal transduction, the mechanisms of switching on/off aflatoxin production, the genes involved in pathogenesis and survival of the fungus in nature remain unknown. We conducted a comprehensive EST sequencing project and cataloged the expressed genes. This is the first step to study fungal biology and genetic regulation at the genomic scale in *A. flavus* for better understanding of the mechanism of aflatoxin formation in the development of new control strategies to eliminate pre-harvest aflatoxin contamination.

2. Materials and methods

2.1. Fungal strain, media and culture conditions

Aspergillus flavus wild type strain NRRL 3357 (ATCC#20026) was selected for making the EST library. This strain produces abundant amounts of aflatoxins B_1 and B_2 and also produces sclerotia under stressed conditions. In order to make the library as representative as possible for gene expression, fungal mycelia were grown under eight media conditions. These conditions were solid wheat bran, liquid glucose minimal salt (GMS), liquid peptone minimal salt (PMS), potato dextrose broth (PDB), solid rice, liquid YES, GMS plus soybean oil and PMS plus soybean oil. Soybean oil was added to promote the expression of genes for lipid metabolism such as lipases [15]. All of these culture conditions were supportive of aflatoxin formation except PMS liquid medium. The liquid cultures were incubated at 30 °C with constant shaking at 150 RPM. The mycelia were harvested by filtration through miracloth at five different time points (18, 24, 36, 48, 72 and 96 h) following inoculation with a conidial suspension. The harvested mycelial samples (wet weight 20 g from wheat bran samples and 10 g from each of the rest of the samples) were mixed and frozen in liquid nitrogen for RNA purification.

2.2. RNA isolation and normalized cDNA library construction

A normalized cDNA expression library was constructed by Incyte Genomics, Inc. (Palo Alto, CA, USA) with the mixed mycelia provided by this USDA laboratory as starting materials. First, a standard cDNA library was constructed. The protocol included poly(A) trimmed, oligo(dT) primed, and 5′ biased random primed reverse-transcription of cDNA (Incyte proprietary, patent pending). Total RNA was purified from the mixed fungal mycelia using a Qiagen RNeasy Plant Mini Kit (Qiagen, Inc., Valencia, CA, USA). Poly(A)$^+$ mRNA isolation from total RNA was performed using Dynabeads® Oligo(dT)$_{25}$ (Dynal Biotech, Lake Success, NY, USA) and Magnetic Particle Concentrator (MPC) (Dynal Biotech). The double stranded cDNA was directionally cloned into pBlueScript (SK$^+$) vector (Stratagene, La Jolla, CA, USA) at *Not*I/*Eco*RI sites with the T7 and T3 promoter sequences in 5′ and 3′ end, respectively. The normalization protocol is a modified version of that described by Soares et al. [21] (proprietary of Incyte Genomics, patent pending) and is referred to as a "Rare Cloned Biased" library. The normalization process involved two rounds of vigorous column hybridizations to remove the abundant copies of transcripts for maximal increase in gene discovery rate. The normalized cDNA library clones were transformed into bacterial *Escherichia coli* DH10B T1 resistant cells. The bacterial colonies (50,000) carrying normalized *A. flavus* cDNA clones were picked onto 384-well format low cross-talk polypropylene NUNC plates (Nanogen, San Diego, CA, USA) for sequencing at The Institute for Genomic Research (TIGR, Rockville, MD, USA).

2.3. Sequencing, annotation and functional Gene Ontology (GO) assignment

Escherichia coli bacterial cells harboring the *A. flavus* cDNA clones were grown in yeast tryptone medium (Biofluids, Rockville, MD, USA) for 18 h at 37 °C with constant shaking at 150 RPM. The cDNA templates were prepared using the Eppendorf-5 Prime Direct Bind prep kit (Eppendorf, Boulder, CO, USA). Single pass, unidirectional (5′ end) sequencing was performed at TIGR on ABI 3700 sequencing machines using standard sequencing methods. Base calling was made using Phred and Trace Tuner (Paracel, Pasadena, CA, USA). The trace file sequences were cleaned using standard TIGR program to trim off vector and adaptor sequences on both 5′ and 3′ ends and to remove low-quality bases. Tentative consensus sequences (TC) were assembled at high stringency from the ESTs sharing overlap regions of greater than 94% identity of over 40 or more continuous bases using the CAP3 program and Paracel Transcript Assembler ([22]; version 2.6.2, http://www.paracel.com) with modifications by the TIGR bioinformatics team. Overlaps based exclusively on low-complexity regions were excluded. Sequences that were not assembled into a TC were termed singleton ESTs. Both TCs and singletons are unique EST sequences. GO term assignments is a specific term used in Gene Index construction for functional classification of genes in standard GO vocabulary using appropriate GO tools. GO provides three structured networks of defined terms to describe gene product attributes (for more information please visit http://www.geneontology.org/GO.doc.html for detail). The

TC and singleton ESTs were searched against a non-redundant protein database to assign a putative function (GO) for each sequence and to construct the *A. flavus* Gene Index at TIGR. The statistical significance threshold for reporting matches against database sequences in the blast search was set at default value 10 (expect value) for the data presented in Tables 2–5.

3. Results

3.1. Sequencing and assembly

A total of 26,110 normalized *A. flavus* cDNA clones were sequenced and 22,037 high quality usable sequences were obtained. After comparison and assembly of overlapping sequences, 7218 unique sequences were identified. These unique sequences consisted of 3749 TC that shared overlapping sequences to other ESTs and 3469 singletons that did not share overlapping sequence. The genes identified in these ESTs account for an estimated 60% of the predicted 12,000 functional genes in the *A. flavus* genome.

3.2. Aspergillus flavus *gene ontology assignments*

The identified unique ESTs (genes) were blasted against a non-redundant protein database. Only about 66% of the 7218 ESTs had homologous counterpart genes in the database. The remaining unique EST sequences (34%) did not have homologous sequences in the existing databases, suggesting that a significant number of the *A. flavus* genes identified in the EST library are novel. These annotated unique EST sequences have been made available to the public at the NCBI GenBank Database (http://www.ncbi.nlm.nih.gov/). The Gene Ontology data were compiled to construct the *A. flavus* gene index that can be accessed and searched at the TIGR web site (http://www.tigr.org/tdb/tgi/) and the *Aspergillus flavus* web site (http://www.aspergillusflavus.org/). The classification of the molecular functions was shown in Table 1.

3.3. Genes of interest identified

Among the ESTs having homologies to the existing GenBank databases, many could be potentially involved directly or indirectly in aflatoxin production such as in global regulation, signal transduction, pathogenicity, virulence, and fungal development. The gene categories and their putative functional classifications are presented in Tables 2–5 below. Note that many more ESTs were found in these categories as listed in Tables 3–5. For simplicity only the top 20 ESTs are listed in each category.

Table 1
Aspergillus flavus gene (EST) ontology assignments

Molecular function	1015 TC/singleton
Enzyme	401
Binding	231
Transporter	110
Structural molecule	74
Molecular function unknown	65
Signal transducer	33
Transcription regulator	24
Translation regulator	23
Obsolete molecular function	18
Chaperone	15
Enzyme regulator	13
Cell adhesion molecule	3
Defense/immunity protein	2
Protein tagging	1
Motor	1
Apoptosis regulator	1
Cellular component	**1378 TC/singleton**
Cell	1278
Cellular component unknown	72
Extracellular	15
Unlocalized	7
Obsolete molecular function	4
Cell wall	2
Biological process	**2376 TC/singleton**
Cell growth and/or maintenance	2090
Cell communication	169
Development	38
Obsolete molecular function	31
Biological process unknown	19
Death	13
Physiological processes	12
Behavior	4

3.3.1. Genes directly involved in aflatoxin biosynthesis

The genes directly involved in aflatoxin formation comprise an aflatoxin pathway gene cluster (25 genes) in *A. parasiticus* and *A. flavus* [6,9]. With only four exceptions [*aflU* (*cypA*), *aflA* (*fas-2*), *aflN* (*verA*) and *aflI* (*avfA*)], all of the aflatoxin pathway genes that were located within the aflatoxin pathway gene cluster in *A. parasiticus* (AY371490) were identified from the *A. flavus* EST database (Table 2). In addition, four new transcripts expressed in the EST library (Table 2, TC4669, NAFAG57TV, TC4876 and TC10997) were identified to have significant homologies to the aflatoxin pathway gene cluster sequence in *A. parasiticus* ([6,9] AY371490). TC 10997 sequence in *A. flavus* corresponding to an ORF encoding for a hypothetical protein consisting of 495 amino acids in *A. parasiticus*. This ORF was named *aflY* (*hypA*) and was reported earlier [6,9]. The relative positions of the three small ORFs were located (Table 2) in the intergenic regions of the reported aflatoxin pathway cluster genes. TC4669, NAFAG57TV and TC4876 are 485, 437 and 491 bp in length and are capable of encoding a

Table 2
Aspergillus flavus ESTs (genes) highly homologous to aflatoxin pathway genes in *A. parasiticus* (AY371490)

A. flavus EST ID	Function	Gene in *A. parasiticus*
TC10184	Dehydrogenase	*aflF* (*norB*)
ni	Cytochrome P450 monooxygenase	*aflU* (*cypA*)
TC11683	Transmembrane protein	*aflT* (*aflT*)
TC10553, TC10554, TC9305	Polyketide synthase	*aflC* (*pksA*)
TC4669	Unknown, hypothetical protein	New ORF
TC6616, TC8487	Norsolorinic acid reductase, ketoreductase	*aflD* (*nor-1*)
ni	Fatty acid synthase alpha subunit	*aflA* (*fas-2*)
NAGAC89TV	Fatty acid synthase beta subunit	*aflB* (*fas-1*)
TC8393, TC10532	Transcription activator	*aflR* (*aflR*)
TC10671	Transcription enhancer	*aflS* (*aflJ*)
TC7096	Alcohol dehydrogenase	*aflH* (*adhA*)
TC10353, NAGAD41TV	Esterase	*aflJ* (*estA*)
TC8710, NAFDF68TV	Norsolorinic acid reductase, dehydrogenase	*aflE* (*norA*)
TC10529	Dehydrogenase/Ketoreductase	*aflM* (*ver-1*)
NAFAG57TV	Unknown, hypothetical protein	New ORF
ni	Monooxygenase	*aflN* (*verA*)
TC10256, TC9785	Cytochrome P450 monooxygenase	*aflG* (*avnA*)
TC9928, NAFEU51TV	Cytochrome P450 monooxygenase, desaturase	*aflL* (*verB*)
TC4876	Unknown, hypothetical protein	New ORF
ni	Averufin oxidase	*aflI* (*avfA*)
TC9044	*O*-methyltransferase B	*aflO* (*omtB*)
TC8404, TC4662, TC10536	*O*-methytransferase A	*aflP* (*omtA*)
TC10424	OMST-oxidoreductase, P450 monooxygenase	*aflQ* (*ordA*)
TC6588, TC11744	VERB synthase	*aflK* (*vbs*)
TC10899	Cytochrome P450 monooxygenase	*aflV* (*cypX*)
TC10309, TC10310	Monooxygenase	*aflW* (*moxY*)
NAGDE85TV	Monooxygenase, oxidase	*aflX* (*ordB*)
TC10997, NAFCZ47TV	Unknown, hypothetical protein	*aflY* (*hypA*)

Note: "ni" in the first column means "not identified"; the ESTs are listed in the same order as the genes in aflatoxin pathway gene cluster in *A. parasiticus*; the new gene names are used [6] and the old names are included in parentheses; the function of the four new ORFs was unknown, hypothetical protein; ORF; open reading frame.

polypeptide of 161, 109 and 163 amino acids, respectively. Though these three ORFs are very small in size and in the intergenic regions, TC4669 and TC4876 were expressed as high as 17 and 7 copies, respectively, even in this normalized cDNA library. We can speculate that these small ORFs could play certain roles in aflatoxin formation. The homologies of aflatoxin pathway genes between *A. flavus* and *A. parasiticus* are extremely high ranging from 90% to 99% with an average of 95% at both nucleotide and amino acid levels. It is quite possible that additional, yet undiscovered genes, involved in aflatoxin synthesis might be located outside of the identified gene cluster.

3.3.2. Genes putatively involved in global regulation and signal transduction

Global regulation of aflatoxin formation is of great interest but is also the least known aspect of aflatoxin biosynthesis. It is well understood that aflatoxin biosynthesis is under the tight control of the positive transcription activator *aflR* gene [10,11,23]. However, the gene or genes that control *aflR* expression are unclear. Among the unique ESTs in the *A. flavus* cDNA library, many of the genes potentially involved in regulation were identified based on sequence comparison against the known genes in the databases (Table 3). Some of these genes are related to stress responses such as mitogen-activated protein kinase (MAPK), MAPK kinase (MAPKK) and MAPKK kinase (MAPKKK) (Campbell, personal communication). These genes could potentially play important roles in signal transduction pathway in response to developmental or environmental elicitors that turn on aflatoxin production. The homolog of a regulatory gene, *laeA* (Accession Nos.: AY394722 in *A. nidulans* and AY422723 in *A. fumigatus*) for loss of *aflR* expression [24], was found (NAGEM53TV) to be expressed in *A. flavus* cDNA library.

3.3.3. Genes possibly involved in virulence/pathogenicity

Invasion of host plants such as corn, cotton, treenuts and peanuts, by *A. flavus* is a complicated process involving many genetic and biological factors. Identification of the genes responsible for such biological processes is a very cumbersome process using conventional molecular cloning methods. Examining the unique ESTs, we identified many hypothetical genes that have the potential to contribute to fungal

Table 3
Genes putatively involved in regulatory/signal transduction

EST ID	Hit accession #	Putative function	Organism	% ID	% sim	*E* value
TC7659	SP|O94321	Multistep phosphorelay regulator 1	*Schizosaccharomyces pombe*	66.67	78.26	1.30E – 26
TC10096	SP|Q03172	Zinc finger protein 40 (transcription factor α)	*Mus musculus*	80.65	87.10	6.70E – 24
TC10270	GB|CAD70954.1	Related to signal recognition particle 72 kDA protein	*Neurospora crassa*	41.98	57.00	2.20E – 47
TC11694	GB|AAK69672.1	myc-type bHLH transcription factor	*Candida albicans*	32.50	54.17	1.10E – 06
TC5059	SP|O13724	Zinc-finger protein zpr1	*Schizosaccharomyces pombe*	50.00	73.00	1.00E – 36
TC11900	SP|P57074	Transcription factor SOX-8	*Gallus gallus*	99.44	100.00	5.20E – 94
TC10472	SP|CAB91681.2	Related to AP-1-like transcription factor	*Neurospora crassa*	44.83	60.69	2.40E – 23
TC9925	GB|CAB91681.2	Related to AP-1-like transcription factor	*Neurospora crassa*	44.05	56.35	1.20E – 37
TC11837	GB|CAB40167.1	Phosphatidylinositol kinase	*Schizosaccharomyces pombe*	67.92	82.08	8.40E – 74
TC11260	GB|AAD40816.1	Histidine kinase	*Nectria haematococca mpVI*	78.71	90.32	6.30E – 66
TC11661	SP|P32586	Protein-tyrosine phosphatase 2	*Schizosaccharomyces pombe*	66.15	76.92	7.80E – 24
TC8756	PIR|H87338	Sensor histidine kinase/response regulator	*Caulobacter crescentus*	39.73	64.38	2.10E – 06
TC9598	GB|BAB47691.1	Transcriptional regulator	*Mesorhizobium loti*	48.24	56.47	5.60E – 10
TC8343	GB|CAD28436.1	Probable osmotic sensitivity map kinase	*Aspergillus fumigatus*	90.41	94.52	1.60E – 67
TC8342	GB|CAD28436.1	Probable osmotic sensitivity map kinase	*Aspergillus fumigatus*	99.11	100.00	8.50E – 55
NAFCH11TV	GB|AAL30826.1	Two-component osmosensing histidine kinase	*Botryotinia fuckeliana*	62.61	81.74	2.70E – 66
TC11834	GB|AAD24428.1	MAP protein kinase (MAPK or MPKA)	*Emericella nidulans*	96.41	98.21	7.00E – 115
TC9026	GB|AAM82166.1	MAP kinase kinase (MAPKK)	*Magnaporthe grisea*	49.22	65.63	5.40E – 19
TC11700	GB|AAL77223.1	Bck1-like MAP kinase kinase kinase (MAPKKK)	*Podospora anserina*	81.31	91.59	4.50E – 91
NAGEM53TV	GI|37622141	Methyltransferase (*laeA*)	*Aspergillus nidulans*	65.00	86.00	3.00

Table 4
Genes putatively involved in virulence/pathogenicity

EST ID	Hit accession #	Putative function	Organism	% ID	% sim	*E* value
NAGBA52TV	GB|AAL30767.1	Parasitic phase-specific protein PSP-1	*Coccidioides posadasii*	39.64	59.76	9.20E – 26
NAGBB06TV	GB|AAF40140.1	β (1–3) glucanosyltransferase Gel3p	*Aspergillus fumigatus*	60.71	75.00	0.026
NAGBM43TV	GB|BAA34996.1	Oligo-1,4-1,4-glucantransferase	*Saccharomyces cerevisiae*	62.13	71.01	3.10E – 61
NAGCR76TV	GB|AAM77702.1	Endoglucanase	*Emericella desertorum*	70.00	86.15	4.90E – 50
TC10675	GB|AAC49904.1	Mixed-linked glucanase precursor	*Cochliobolus carbonum*	86.21	91.30	1/E – 48
TC11507	PIR|S59841	4-α-Glucanotransferase/amylo-1,6-glucosidase	*Saccharomyces cerevisiae*	62.60	77.24	7.30E – 60
TC11738	GB|CAD24293.1	β-Galactosidase	*Aspergillus candidus*	99.54	100.00	5.90E – 118
TC11835	GB|BAC07256.1	Cellobiohydrolase D	*Aspergillus oryzae*	100.00	100.00	3/E – 64
TC11846	GB|AAL84695.1	β-1,3-Glucanase precursor	*Hypocrea virens*	48.34	60.93	3.10E – 30
TC8364	GB|AAL09828.1	β-Glucosidase 4	*Coccidioides posadasii*	61.92	75.50	4.80E – 100
TC8391	PIR|A25494	Hydroxyproline-rich glycoprotein	*Lycopersicon esculentum*	33.33	48.61	0.0047
TC8585	GB|AAF40140.1	β (1–3) Glucanosyltransferase Gel3p	*Aspergillus fumigatus*	77.55	83.67	2.40E – 59
TC8959	GB|AAK58059.1	Glucan 1,3 β-glucosidase-like protein	*Ophiostoma novo-ulmi*	70.83	81.25	4/E – 50
TC9038	GB|AAD43340.1	Pectin methylesterase	*Cochliobolus carbonum*	34.64	50.98	2.10E – 12
TC9071	PIR|T11674	Glutamine-fructose-6-phosphate transaminase	*Schizosaccharomyces pombe*	61.39	75.95	3.90E – 50
TC9141	GB|AAD01641.1	Pathogenicity protein	*Magnaporthe grisea*	45.45	68.18	1.40E – 28
TC9271	GB|BAB69770.1	Glycogen branching enzyme	*Aspergillus oryzae*	100.00	100.00	5.30E – 94
TC10631	GB|AAL95714.1	Antigenic cell wall protein MP1	*Aspergillus flavus*	100.00	100.00	7.00E – 131
NAGBG48TV	GB|BAA35140.1	Chitinase	*Emericella nidulans*	68.55	84.68	1.30E – 4
TC9836	SP|Q90121	Killer protein 4 (KP4) toxin precursor (fungal toxin)	*Ustilago maydis virus P4*	30.95	47.62	1.70E – 05

Table 5
Genes putatively involved in stress response and antioxidation

EST ID	Hit accession #	Putative function	Organism	% ID	% sim	*E* value
NAFAE55TH	GB|AAM73769.1	Stress response element binding protein	*Trichoderma atroviride*	59.63	72.67	7.90E − 43
NAFBN21TV	GB|AAK54753.1	Thiol-specific antioxidant	*Ajellomyces capsulatus*	70.97	83.87	3.70E − 68
NAGAG45TV	GB|AAK54753.1	Thiol-specific antioxidant	*Ajellomyces capsulatus*	77.12	87.29	6.10E − 45
NAGCF11TV	GB|CAA60962.1	Oxidative stress resistance	*Saccharomyces cerevisiae*	55.00	73.33	1.90E − 11
TC8386	GB|BAC56176.1	Cu,Zn superoxide dismutase	*Aspergillus oryzae*	100.00	100.00	8.30E − 80
TC10087	GB|AAK17008.1	Mn-superoxide dismutase	*Emericella nidulans*	82.89	88.60	9.00E − 99
TC10360	SP|O43099	Probable peroxisomal membrane protein PMP20	*Aspergillus fumigatus*	86.31	92.26	6.90E − 76
TC10820	SP|O43099	Probable peroxisomal membrane protein PMP20	*Aspergillus fumigatus*	68.48	83.03	1.80E − 59
TC10350	SP|Q9UW02	Thioredoxin (Allergen Cop c 2)	*Coprinus comatus*	51.55	70.10	1/E − 23
TC10342	SP|P29429	Thioredoxin	*Emericella nidulans*	64.42	82.69	2.50E − 32
TC9135	PIR|T48748	Probable glutaredoxin 8D4.220	*Neurospora crassa*	60.38	79.25	2/E − 27
TC10000	GB|AAQ84041.1	Peroxisomal-like protein	*Paracoccidioides brasiliensis*	61.02	83.05	9.80E − 42
NAGAY31TV	GB|AAH58481.1	Peroxiredoxin 2	*Rattus norvegicus*	100.00	100.00	1.70E − 104
NAGER47TV	SP|P34723	Thioredoxin	*Penicillium chrysogenum*	57.94	77.57	1.30E − 28
TC9574	PIR|T51908	Glutathione-disulfide reductase	*Neurospora crassa*	66.18	82.84	9.40E − 72
TC10693	GB|CAB56542.1	Zinc/cadmium resistance protein	*Saccharomyces cerevisiae*	49.39	70.12	2.90E − 38
NAFFI48TV	SP|Q08220	Glutathione synthetase	*Saccharomyces cerevisiae*	44.30	60.53	5.10E − 41
NAFDF35TV	SP|Q8X0X0	Glutamate–cysteine ligase	*Neurospora crassa*	61.69	75.62	9.20E − 58
NAFFN33TV	GB|AAL13750.1	LD22804p	*Drosophila melanogaster*	36.21	61.49	2.70E − 24
TC9435	GB|AAG45152.2	Catalase C	*Emericella nidulans*	45.74	65.89	1.20E − 24

virulence or pathogenicity (Table 4). These genes encode hydrolytic enzymes, which could be highly expressed virulence factors during fungal invasion of *A. flavus* into crop plants.

3.3.4. Genes possibly involved in stress response and antioxidation

Jayashree, T. and Subramanyam [25] reported that oxidative stress triggered aflatoxin biosynthesis in *A. parasiticus* [25]. On the other hand hydrolyzable tannins acting as antioxidants in living cells [26] completely arrested aflatoxin biosynthesis while having little effect on fungal growth [27]. The active anti-aflatoxigenic constituent of these tannins was identified as gallic acid. In the *A. flavus* EST database, over three dozen genes hypothetically involved in stress responses and anti-oxidation were identified (Table 5). By comparison to the *A. flavus* ESTs the oxidative stress response pathway gene homologs were also identified in *Saccharomyces cerevisiae* and their functionality determined in yeast deletion mutant strains (Campbell, unpublished data).

3.3.5. Genes putatively involved in fungal development and sporulation

Secondary metabolism is often correlates with fungal developmental processes such as sporulation and sclerotia formation [16,18,28]. Mutants that are deficient in sporulation are unable to produce aflatoxins [18,28]. A critical advance in this regard was the finding that the regulation of sporulation and ST production is by means of a shared Gprotein mediated growth pathway in *A. nidulans* [19,29]. Several genes involved in fungal development and conidiation were identified in the *A. flavus* EST library (NAGBA10TV, NAGBD49TV, NAGBU14TV, NAFEA74TV, NAGBI96TV, TC10246, TC11956, TC12083 and TC8878).

4. Discussion

Applying an EST strategy provided a rapid and effective method for identification of genes potentially involved in aflatoxin contamination of crops by *A. flavus*. We reported the identification of 7218 unique genes in this *A. flavus* cDNA expression library. The average cDNA insert size of the library is about 1.2–1.5 kb with high quality sequences. However, due to the fact that 34% of the unique genes do not have homologs identified in the database, there is a possibility that two or more non-overlapping ESTs may be transcribed from the same gene and were counted as two or more unique genes. Further, oligo-dT primed cDNA library is often notorious for not being full length (5′ truncation). Over-estimation of unique genes based purely on bioinformatics information is unavoidable. Therefore, the

actual number of unique genes could be a little less than 7218. There were four aflatoxin pathway genes that were not identified in this library. The EST copy number of a specific gene (in TC group) does not truly reflect the level of gene expression in a normalized library like this one. Our experience on secondary metabolism gene expression, such as that required for aflatoxin biosynthesis, indicated that the aflatoxin pathway genes are expressed at a much lower level than primary metabolism pathway genes. In addition, we identified only over 7218 unique genes from over 26,000 cDNA clones sequenced, which account for 60% of the total functional genes in the fungal genome. Considering that 4 out of 25 aflatoxin pathway genes were not identified in the library, it is consistent with the overall probability. The genes identified in this study are the putative candidates for further investigation. Genes responsible for the biosynthesis of secondary metabolites such as aflatoxins are those encoding polyketide synthases, fatty acid synthases, carboxylases, dehydrogenases, reductases, oxidases, oxidoreductases, epoxide hydrolases, mono- or di-oxigenases, cytochrome P450 monooxigenases, and methyltransferases [6,14]. In the *A. flavus* EST database, numerous genes fall within the categories of these enzymes. Without additional biological evidence it is very difficult to predict whether these genes are involved in primary or secondary metabolisms based purely on the bioinformatic annotations. Also, as mentioned earlier, about 34% of the ESTs do not have homologs identified in the existing databases. In order to identify the most prominent candidate genes, such as global regulators in *A. flavus*, a comprehensive screening or profiling of those genes requires additional genomic scale studies such as gene expression profiling by microarray experiments followed by analysis of targeted mutagenesis.

Acknowledgement

The authors express their sincere thanks to Catherine Ronning, The Institute for Genomic Research, for her help in bioinformatic analysis and data organization. We are grateful for the helpful suggestions and positive input for the *Aspergillus flavus* EST project and for preparation of this manuscript from Prof. Gary A. Payne and Prof. Ralph A. Dean, North Carolina State University, from Prof. Joan W. Bennett, Tulane University, Keietsu Abe, Tohoku University, Japan, and Dr. Masayuki Machida, National Institute of Advanced Industrial Science and Technology (AIST), Japan. We thank Dr. Campbell for sharing unpublished data. We thank Dr. Keller for sharing *A. nidulans laeA* sequence information. We are thankful for the secretarial help of Janell Becker.

References

[1] Cotty, P.J. (1997) Aflatoxin-producing potential of communities of *Aspergillus* section Flavi from cotton producing areas in the United States. Mycol. Res. 101, 698–704.
[2] Bennett, J.W. and Klich, M. (2003) Mycotoxins. Clin. Microbiol. Rev. 16, 497–516.
[3] Lancaster, M.D., Jenkins, F.P. and Phillip, J.M. (1961) Toxicity associated with certain samples of groundnuts. Nature 192, 1095–1096.
[4] Bressac, B., Kew, M., Wands, J. and Ozturk., M. (1991) Selective G to T mutations of p53 gene in hepatocellular carcinoma from Southern Africa. Nature 350, 429–431.
[5] Bhatnagar, D., Ehrlich, K.C. and Cleveland, T.E. (2003) Molecular genetic analysis and regulation of aflatoxin biosynthesis. Appl. Microbiol. Biotechnol. 61, 83–93.
[6] Yu, J., Chang, P.-K., Ehrlich, K.C., Cary, J.W., Bhatnagar, D., Cleveland, T.E., Payne, G.A., Linz, J.E., Woloshuk, C.P. and Bennett, J.W. (2004) The clustered pathway genes in aflatoxin biosynthesis. Appl. Environ. Microbiol. 70, 1253–1262.
[7] Yu, J., Proctor, R.H., Brown, D.W., Abe, K., Gomi, K., Machida, M., Hasegawa, F., Nierman, W.C., Bhatnagar, D. and Cleveland, T.E. (2004) Genomics of economically significant *Aspergillus* and *Fusarium* species In: Applied Mycology and Biotechnology (Arora, K.D., Khachatourians, G.G., Eds.), Fungal Genomics, Vol. 4, pp. 249–283. Elsevier, Amsterdam.
[8] Yu, J., Chang, P.-K., Cary, J.W., Wright, M., Bhatnagar, D., Cleveland, T.E., Payne, G.A. and Linz, J.E. (1995) Comparative mapping of aflatoxin pathway gene clusters in *Aspergillus parasiticus* and *Aspergillus flavus*. Appl. Environ. Microbiol. 61, 2365–2371.
[9] Yu, J., Bhatnagar, D. and Cleveland, T.E. (2004) Completed sequence of aflatoxin pathway gene cluster in *Aspergillus parasiticus*. FEBS Lett. 564, 126–130.
[10] Chang, P.-K., Cary, J.W., Bhatnagar, D., Cleveland, T.E., Bennett, J.W., Linz, J.E., Woloshuk, C.P. and Payne, G.A. (1993) Cloning of the *Aspergillus parasiticus apa-2* gene associated with the regulation of aflatoxin biosynthesis. Appl. Environ. Microbiol. 59, 3273–3279.
[11] Payne, G.A., Nystrom, G.J., Bhatnagar, D., Cleveland, T.E. and Woloshuk, C.P. (1993) Cloning of the *afl-2* gene involved in aflatoxin biosynthesis from *Aspergillus flavus*. Appl. Environ. Microbiol. 59, 156–162.
[12] Meyers, D.M., O'Brian, G., Du, W.L., Bhatnagar, D. and Payne, G.A. (1998) Characterization of *afl*J, a gene required for conversion of pathway intermediates to aflatoxin. Appl. Environ. Microbiol. 64, 3713–3717.
[13] Chang, P.-K. (2003) The *Aspergillus parasiticus* protein *AFLJ* interacts with the aflatoxin pathway-specific regulator AFLR. Mol. Genet. Genomics 268, 711–719.
[14] Payne, G.A. and Brown, M.P. (1998) Genetics and physiology of aflatoxin biosynthesis. Annu. Rev. Phytopathol. 36, 329–362.
[15] Yu, J., Mohawed, S.M., Bhatnagar, D. and Cleveland, T.E. (2003) Substrate-induced lipase gene expression and aflatoxin production in *Aspergillus parasiticus* and *Aspergillus flavus*. J. Appl. Microbiol. 95, 1334–1342.
[16] Calvo, A.M., Wilson, R.A., Bok, J.W. and Keller, N.P. (2002) Relationship between secondary metabolism and fungal development. Microbiol. Mol. Biol. Rev. 66, 447–459.
[17] Chang, P.-K., Bennett, J.W. and Cotty, P.J. (2001) Association of aflatoxin biosynthesis and sclerotial development in *Aspergillus parasiticus*. Mycopathologia 153, 41–48.
[18] Kale, S.P., Cary, J.W., Bhatnagar, D. and Bennett, J.W. (1996) Characterization of experimentally induced, nonaflatoxigenic variant strains of *Aspergillus parasiticus*. Appl. Environ. Microbiol. 62, 3399–3404.

[19] Hicks, J.K., Yu, J.H., Keller, N.P. and Adams, T.H. (1997) *Aspergillus* sporulation and mycotoxin production both require inactivation of the FadA G alpha protein-dependent signaling pathway. EMBO J. 16, 4916–4923.
[20] Payne, G.A. (1998) Process of contamination by aflatoxin-producing fungi and their impact on crops In: Mycotoxins in Agriculture and Food Safety (Sinha, K.K., Bhatnagar, D., Eds.), pp. 279–306. Marcel Dekker, New York, NY.
[21] Soares, M.B., Bonaldo, M.F., Jelene, P., Su, L., Lawton, L. and Efstratiadis, A. (1994) Construction and characterization of a normalized cDNA library. Proc. Natl. Acad. Sci. USA 91, 9228–9232.
[22] Huang, X. and Madan, A. (1999) CAP3: A DNA sequence assembly program. Genome Res. 9, 868–877.
[23] Ehrlich, K.C., Montalbano, B.G. and Cary, J.W. (1999) Binding of the C6-zinc cluster protein, AFLR, to the promoters of aflatoxin pathway biosynthesis genes in *Aspergillus parasiticus*. Gene 230, 249–257.
[24] Bok, J.W. and Keller, N.P. (2004) LaeA, a regulator of secondary metabolism in *Aspergillus* spp. Eukaryot. Cell 3, 527–535.
[25] Jayashree, T. and Subramanyam, C. (2000) Oxidative stress as a prerequisite for aflatoxin production by *Aspergillus parasiticus*. Free Radic. Biol. Med. 29, 981–985.
[26] Sroka, Z. and Cisowski, W. (2003) Hydrogen peroxide scavenging, antioxidant and anti-radical activity of some phenolic acids. Food Chem. Toxicol. 41, 753–758.
[27] Mahoney, N. and Molyneux, R.J. (2004) Phytochemical inhibition of aflatoxigenicity in *Aspergillus flavus* by constituents of walnut (*Juglans regia*). J. Agric. Food Chem. 52, 1882–1889.
[28] Bennett, J.W., Leong, P.M., Kruger, S. and Keyes, D. (1986) Sclerotial and low aflatoxigenic morphological variants from haploid and diploid *Aspergillus parasiticus*. Experientia 42, 848–851.
[29] Yu, J.H., Weiser, J. and Adams, T.H. (1996) The *Aspergillus* FlbA RGS domain protein antagonizes G-protein signaling to block proliferation and allow development. EMBO J. 15, 5184–5190.